0273-2289/05/121–124/129–148/$30.00

Catalytically Enhanced Endocellulase Cel5A from *Acidothermus cellulolyticus*

John O. Baker,*[,1] James R. McCarley,[2] Rebecca Lovett,[2] Ching-Hsing Yu,[2] William S. Adney,[1] Tauna R. Rignall,[1] Todd B. Vinzant,[1] Stephen R. Decker,[1] Joshua Sakon,[2] and Michael E. Himmel[1]

[1]National Bioenergy Center, National Renewable Energy Laboratory, 1617 Cole Boulevard, Golden, CO 80401, E-mail: john_baker@nrel.gov; and [2]Department of Chemistry and Biochemistry, University of Arkansas, Fayetteville, AR 72701

Abstract

When Tyr245 in endocellulase Cel5A from *Acidothermus cellulolyticus* was changed to Gly (Y245G) by designed mutation, the value of K_i for inhibition of the enzyme by the product cellobiose was increased more than 1480%. This reduction in product inhibition enabled the mutant enzyme (used in conjunction with *Trichoderma reesei* cellobiohydrolase-I) to release soluble sugars from biomass cellulose at a rate as much as 40% greater than that achieved by the wild-type (WT) enzyme. The mutant was designed on the basis of the previously published crystal structure of the WT enzyme/substrate complex (at a resolution of 2.4 Å), which provided insights into the enzyme mechanism at the atomic level and identified Tyr245 as a key residue interacting with a leaving group. To determine the origin of the change in activity, the crystal structure of Y245G was solved at 2.4-Å resolution to an *R*-factor of 0.19 (*R*-free = 0.25). To obtain additional information on the enzyme-product interactions, density functional calculations were performed on representative fragments of the WT Cel5A and Y245G. The combined results indicate that the loss of the platform (Y245G) and of a hydrogen bond (from a conformational change in Gln247) reduces the binding energy between product and enzyme by several kilo calories per mole. Both kinetic and structural analyses thus relate the increased enzymatic activity to reduced product inhibition.

Index Entries: Bioethanol; biomass conversion; protein crystallography; endoglucanase; *Acidothermus cellulolyticus*; *Trichoderma reesei*.

Introduction

Cellulose, an unbranched β-1,4-linked homopolymer of glucose, is the most abundant renewable fuel resource on Earth, accounting for about

*Author to whom all correspondence and reprint requests should be addressed.

half of the organic material in the biosphere, and is the major polysaccharide found in plant biomass. The hydrolysis of cellulose, aided by endocellulase, exocellulase, and β-D-glucosidase catalysis, produces glucose, an easily fermentable monosaccharide. Intense research is aimed at the conversion of cellulose into ethanol because, as a source of renewable fuel, the process has great economic potential and is environmentally friendly *(1–3)*. Unfortunately, a major barrier to the efficient production of ethanol from biomass cellulose is the low activity of cellulases against crystalline cellulose. Therefore, our main objective is to choose the best enzymes and improve their activities further.

Cellulose is insoluble and crystalline and, hence, largely resistant to microbial attack. In many biomass utilization schemes, the raw material is first treated with dilute acid at moderate temperatures *(4)* to remove hemicellulose and to speed up cellulose hydrolysis. The pretreated biomass can then be subjected to carefully chosen mixtures of endo- and exoglucanases for maximum cost-effectiveness *(5)*. The *Acidothermus cellulolyticus* Cel5A endoglucanase is one of the most active cellulases known *(6)*. In combination with the exocellulase 1,4-β-cellobiohydrolase (CBH I) from *Trichoderma reesei (7)*, Cel5A gives the highest saccharification and degree of synergism of all cellulases tested *(8)*.

In addition to a catalytic domain, most cellulases possess one or two cellulose-binding domains (CBD). Even though the presence of a CBD is essential for the degradation of cellulose by exoglucanases, it is not required by many endoglucanases *(9)*. In the current study, we wished to explore whether protein active-site engineering can be used to enhance the activity of cellulases. Specifically, we wished to increase the activity by engineering mutants that speed up reaction owing to reduced product-enzyme interactions. To guide protein engineering by site-directed mutagenesis, we solved the crystal structure of the catalytic domain of Cel5A in complex with cellotetraose (Fig. 1) *(10)*. Cel5A is a member of the 4/7 superfamily or Clan A of glycosyl hydrolases (CAZy Web site at http://afmb.cnrs-mrs.fr/CAZY/) and, as such, is a retaining enzyme and has an $(\alpha/\beta)_8$-barrel structure. The generally accepted catalytic mechanism of these superfamily members is a double-displacement mechanism that was originally proposed by Koshland *(11)*. That mechanism involves an initial binding of the substrate to the enzyme followed by a general acid-catalyzed attack of an enzymatic nucleophile (Glu282) on the anomeric carbon that proceeds through an ionic transition state to form a glycosyl-enzyme intermediate *(10,12,13)*. This intermediate is then hydrolyzed by a general base-catalyzed attack of water on the anomeric carbon that, again, proceeds through an ionic transition state, forming the product and returning the enzyme to its original protonation state *(12,13)*. The cellotetraose molecule is bound in a manner consistent with the expected Michaelis complex for the glycosylation half-reaction. The structure yielded detailed information on the enzyme-substrate interactions that made it possible to

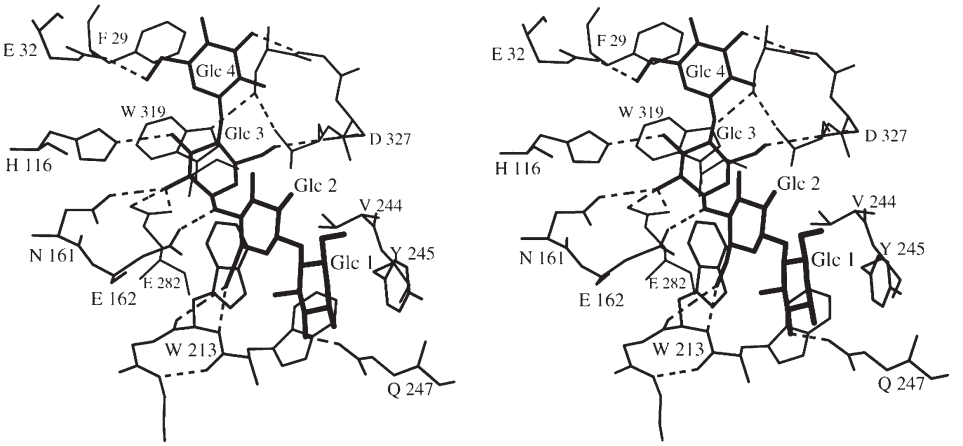


Fig. 1. Stereoscopic view of active site of Cel5A. Sakon et al. *(10)* provide a detailed description of substrate-enzyme interactions.

propose a detailed sequence of physical steps of the cleavage mechanism for Cel5A and to design a catalytically enhanced endocellulase.

The enzyme interacts with all four residues of cellotetraose by both hydrophobic contacts and hydrogen bonding. As is commonly found in polysaccharide-binding enzymes (*14–16*), the hydrophobic face of each glucose unit interacts with an aromatic side chain at the active-site cleft. We showed *(10)* that Glc1, Glc2, Glc3, and Glc4 interact with Tyr245, Trp213, Trp319, and Phe29, respectively. Based on our proposed catalytic mechanism, the leaving group of the glycosylation half-reaction binds to platform residues Tyr245 and Trp213. By mutation of the platforms, it should be possible to make the leaving group bind less tightly. In theory, this decrease in the strength of product binding could increase the rate of the enzyme reaction in at least three different ways, which are described with reference to the scheme of Fig. 2. First, the rate of release of just formed product from the active site could be increased (increase in k_5). The acceleration of this elementary step would result in a measurable increase in the overall rate only if the dissociation of the leaving group is at least partly rate limiting. Second, at the same time, the weakening of the complex between one of the products and the enzyme should slow the reverse reaction or transglycosylation *(17)*, which was observed, e.g., for a 0.3 *M* solution of the product, cellobiose, in both the crystal and in solution *(10)* (in the scheme of Fig. 2, reduction in the concentration of EP and thus in the value of k_4[EP]). Third, for assay conditions such that significant concentrations of soluble products (e.g., cellobiose) are accumulated in the bulk solution, loosening of binding of glucosyl residues at the leaving-group site might be expected to reduce the extent of inhibition by product (Fig. 2, K_i). Between the two platform residues, Tyr245 is farther away from the scissile glycosidic bond than Trp213, so a mutation of the former is less likely to interfere with the cleavage mechanism than

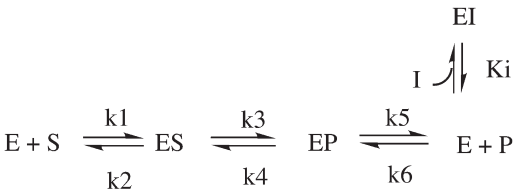


Fig. 2. Proposed scheme for interactions among Cel5A, substrate cellulose, and the principal observed product, cellobiose, referred to here as either P, to denote its role as leaving group, or I, to denote its role as an inhibitor returning from the bulk solution.

a mutation of the latter. For that reason, Tyr245 was chosen as a target for mutagenesis. The mutant was initially tested for hydrolysis of both 4-methylumbelliferyl-β-D-cellobioside (MUC) and highly amorphous cellulose (phosphoric acid–swollen Avicel). It was then tested in conjunction with *T. reesei* cellobiohydrolase (CBH) (Cel7A) for degradation of the ultimate target, dilute-acid-treated hardwood. Amorphous cellulose is an excellent substrate for endoglucanases acting alone, but acid-treated yellow poplar (*Liriodendron tulipifera*) is not. The synergistic action of both endocellulase and CBH (exocellulase) is required for effective degradation of the hardwood.

Structural consequences of the mutation were determined by X-ray crystallography, using crystals grown in the presence of substrates and products. To support our views of the energetics of the enzyme-substrate interactions, density functional (DFT) calculations *(18)* were performed at the heptapeptide (residues 241–247) and a single glucosyl unit level. We demonstrate in this article that structure-guided protein engineering is an effective approach to enhance enzymatic reactions that require two synergistic enzymes *(19)* and are difficult to assay *(20)*.

Materials and Methods

Polymerase Chain Reaction Mutation

A 2.3-kb *Bam*HI fragment, obtained from a λ phage library and carrying most of the *Cel5A* gene, including its native promoter, which functions in both *Escherichia coli* and *Streptomyces lividans*, plus approx 800 bp of upstream sequence, was subcloned into pUC19 *(21).* The downstream *Bam*HI site cleaves the Cel5A coding sequence at a point such that the protein is genetically truncated near the beginning of the cellulose-binding module (CBM).The gene was additionally truncated by the addition of two stop codons (TAATGA) directly after the coding sequence for the catalytic domain. Thus, the construct encodes a protein that includes a signal peptide, the N-terminal catalytic domain without the linker or the partial CBM. A QuickChange SDM kit (Stratagene, San Diego, CA) was used to generate mutants. Template DNA from *E. coli* XL1-blue cells, trans-

formed with Dpn1-treated mutated DNA, was prepared for sequencing using a QIAprep-spin plasmid purification miniprep procedure (Qiagen, New Haven, CT). The transformed XL1-blue cells were grown overnight in 5 mL of Luria-Bertani (LB) broth with 100 μg/mL of ampicillin selection. Cells were removed by centrifugation, and the plasmid was isolated using the protocol outlined in the QIAprep-spin handbook. Following isolation of the plasmid, the presence of the 2.3-kb insert was confirmed by digestion with *Bam*HI followed by agarose electrophoresis. DNA from transformants containing the insert was ethanol precipitated and then polyethylene glycol precipitated. The sequence was determined by the DNA Sequencing Facility at Iowa State University, in order to confirm mutations. To produce the modified proteins, transformed *E. coli* XL1/blue cells were first grown overnight on LB plates with 100 μg/mL of ampicillin at 37°C. A single colony was then used to inoculate a 500-mL baffled Erlenmeyer flask containing 200 mL of LB broth and 100 μg/mL of ampicillin. The seed culture was allowed to grow for 16–20 h at 37°C with shaking at 250 rpm in a reciprocating shaking incubator (New Brunswick Scientific, Edison, NJ). The seed culture was then aseptically transferred to a 10-L BioFlow 3000 fermentor (New Brunswick Scientific) with a working volume of 9 L, in which pH, temperature, agitation, and dissolved oxygen (DO) were held constant.

Fermentation medium consisted of LB broth, 100 μg/mL of ampicillin, and 2.5% filter-sterilized glucose added after autoclaving. The DO polarographic probe was calibrated using nitrogen at 4.0 L/min (for zero) and house air at 4.0 L/min (for 100%). Pure oxygen was blended with air to maintain a constant 20% DO at a flow of 4 L/min. The pH was held at 6.8 with the addition of 2 *M* KOH throughout the fermentation. The temperature was held at 30°C to prevent the formation of inclusion bodies, and the agitation was held at 250 rpm. The fermentations were allowed to grow for 24–28 h, at which time maximum cell densities of 15–20 OD_{600} were usually observed. The cells were harvested using a bench model (LE) CEPA continuous centrifuge (New Brunswick Scientific) at 25,000 rpm (17,000*g*).

Cell Lysis by Bead Milling

Fifty grams of cells (wet weight) was added to the chamber of a Biospec (Bartlesville, OK) stainless steel bead mill containing 200 g of 0.1-mm glass beads and 200 mL of 20 m*M* Tris, pH 8.0. The cells were lysed using the bead mill for 5 min while the chamber was chilled with ice. The contents of the chamber were diluted twofold with buffer and divided into 250-mL centrifuge bottles. Cells and cell debris were removed by centrifuging at 27,500*g* in a Sorvall RC2B centrifuge (Kendro, Asheville, NC) at 13,000 rpm for 25 min in a GSA rotor at 4°C. The supernatant was decanted, the pellet was resuspended in buffer, and the cells were milled and centrifuged again.

Purification of Mutant Cel5A Enzymes from E. coli XL1/Blue Cell Lysate

Two different procedures were evaluated for use in preparing the combined supernatants for chromatographic purification. The first approach involved a heat treatment intended to rid the desired, very thermostable Cel5A of most contaminating *E. coli* proteins by selectively denaturing and precipitating the latter. The pooled supernatant from the cell lysate was brought to 0.5 *M* in ammonium sulfate. The supernatant was then divided into 250-mL centrifuge bottles and heated in a 65°C water bath for 50 min to denature non-Cel5A protein. Precipitated proteins were removed by centrifuging at 27,500*g* in a Sorvall RC2B centrifuge (Kendro) at 13,000 rpm for 25 min in a GSA rotor at 4°C. The supernatant was then filtered through a glass fiber filter pad before chromatography, which is described below.

The second procedure, which resulted in improved Cel5A protein yields, eliminated the need for clarification of the supernatant after lysing the cells. With this procedure the cell lysate, which had been adjusted to a conductivity of <2000 mS/cm, was loaded directly onto a Pharmacia Streamline column packed with DEAE (fluidized at an upflow rate of 15 mL/min with 20 m*M* Tris, pH 8.0 buffer). After the column matrix was washed free of the cell debris and the ultraviolet absorbance returned close to zero, the flow was reversed to a downflow orientation and the proteins were eluted using a linear gradient of 20 m*M* Tris, 1 *M* NaCl, pH 8.0 buffer. Fractions that contained activity were pooled and ammonium sulfate was added to a final concentration of 0.5 *M*. These samples, or the cell lysate treated by the first (heat/centrifugation/filtration) procedure, were then loaded on a Phenyl Sepharose HiLoad column. After the column was washed with 3–5 column vol of the starting buffer, the Cel5A enzyme was eluted by a 3.2-column vol descending linear gradient to zero percent salt in 20 m*M* Tris, pH 8.0. The final purification step and buffer exchange were done using a Superdex 200, HiLoad prep-grade column with a flow rate of 0.5 mL/min in 20 m*M* acetate, pH 5.0 buffer with 100 m*M* NaCl. Mutant enzymes eluted as single, symmetrical peaks, indicating a high level of homogeneity. Protein concentrations were determined spectrophotometrically.

Immunoblots and Western Blots

Solid-phase immunologic techniques were used for the detection of expressed enzyme protein. To verify the presence of Cel5A and Cel5A mutant enzymes, Western and immunoblots blots were performed. For Western blots, 3–5 μg of protein was added to each lane and the proteins were electrophoresed. For immunoblots, 2 μL of chromatography sample was applied to nitrocellulose and allowed to air-dry, after which a monoclonal antibody specific for Cel5A was added. This was followed by the

addition of a goat–antimouse IgG alkaline phosphatase–labeled antibody. The bound Cel5A protein was visualized by the precipitation of the substrate 5-bromo-4-chloro-3-indolyl phosphate/Nitro blue tetrazolium (BCIP/NBT).

Closed-Tube Assay of Native and Mutant Cel5A Enzymes Against Amorphous Cellulose

The principle comparisons between activity of the native and mutant enzymes were based on measurement of the release of soluble sugar from phosphoric acid–swollen cellulose (PASC). Substrate for these assays was prepared from Avicel Type PH 101 (FMC, Philadelphia, PA). Five grams of Avicel was first moistened with deionized water; then 150 mL of 85% phosphoric acid was added slowly over a period of 1 h with gentle stirring, the slurry being maintained in an ice bath. After the addition of 100 mL of cold acetone, the slurry was centrifuged for 10 min at 5000*g*. The pellet was washed three times by resuspension and centrifugation in deionized water and then thoroughly dispersed by blending with an IKA-Werk Ultra-Turrax (Tekmar, Cincinnati, OH). The swollen cellulose was stored at 4°C under acidic conditions (pH <2.0). For assays under the conditions of this series, a 1.25% (w/v) slurry was made and exchanged into the assay buffer by successive centrifugations and resuspensions in 20 m*M* acetate, pH 5.0, 0.02% sodium azide, until the supernatant pH was within 0.01 units of 5.0. Assays were carried out in sealed 1.8-mL high-performance liquid chromatography (HPLC) sample vials at 40°C and pH 5.0 in 20 m*M* acetate. Each 1.0-mL reaction contained 5 mg of acid-swollen cellulose and 28 μg (approx 70 nmol) of either native Cel5A or Y245G mutant. Purified *Aspergillus niger* β-D-glucosidase, where present, was added at 45 μg/mL.

Enzyme or enzyme mixture (total volume of 0.6 mL) was prepared beforehand in each reaction vial; reactions were then initiated by adding 0.4 mL of 1.25% (w/v) PASC in reaction buffer (pipetted from a well-stirred slurry). The reaction vials were sealed with aluminum crimp-caps (silicone septum) and incubated for 4 h in a water bath at 40°C with constant mixing by inversion at 10 rpm (axis of rotation 35° from horizontal, with long dimensions of vials tangential to the direction of rotation). At the end of the incubation period, the vials were removed from the water bath and cooled in an ice bath, and the reaction mixture was quickly removed by a syringe needle through the septum and filtered (Acrodisc-13, 0.2-μm pore size; Pall-Gelman, Ann Arbor, MI) to remove the remaining substrate. Glucose and cellobiose were then measured by chromatography on an HPX-87H column (Bio-Rad, Hercules, CA) at 65°C with 0.01 *N* sulfuric acid (0.6 mL/min) as the mobile phase in an HP 1100 chromatograph with a refractive index detector (HP 1047), both from Agilent (Wilmington, DE).

Diafiltration Saccharification Assay of Native and Mutant Cel5A Enzymes with T. reesei Cel7A

In the diafiltration saccharification assay (DSA) previously developed *(20)*, cellulase enzymes carry out substantial conversion of an insoluble cellulosic substrate in a continuously buffer-swept stirred-tank membrane reactor, with the solubilized saccharification products being continuously removed by the buffer flux through an ultrafiltration membrane (Biomax-5; Millipore), while the insoluble substrate and macromolecular enzymes are retained. The course of the reaction is then followed by HPLC analysis of the soluble products in the effluent stream. All DSAs were carried out at 40°C in 20 m*M*, pH 5.0 sodium acetate buffer (plus 0.02% sodium azide to prevent microbial growth). Substrate loading for each assay was 96 mg (dry weight) of pretreated yellow poplar loaded at 4.36% in terms of biomass and 2.5% in cellulose. The substrate was ground, with the bulk of the material between 10 and 500 μ in maximum dimension.

Enzyme loadings for DSA were constant at 27.7 nanomol of a given wild-type (WT) or mutant *A. cellulolyticus* Cel5A g^{-1} cellulose, plus, in all cases, 526 nanomol of *T. reesei* Cel7A g^{-1} cellulose, for an enzyme cocktail that was 5% endoglucanase and 95% CBH on a molar basis. This proportion of endoglucanase in the binary mixture was chosen to be high enough to give the mixture readily measurable activity, but sufficiently below the optimal endoglucanase proportions (i.e., for sugar release and synergism) for maximal sensitivity to differences in endoglucanase activity.

Inhibition Constants for Cellobiose with WT and Y245G Mutant

Inhibition constants (K_i) for inhibition of the hydrolysis of MUC (Sigma, St. Louis, MO) were determined under conditions matching those of the DSA and closed-tube (PASC) experiments. The enzymes (0.682 ng) were incubated for 30 min with the substrate at each of two concentrations (4 and 20 μ*M*) in the presence of D-cellobiose (Sigma) concentrations ranging from 0 to 5 m*M* for the WT catalytic domain, and from 0 to 50 m*M* for the Y245G mutant. At the end of the incubation period, the reaction in each 1-mL assay mixture was terminated by the addition of 2 mL of 0.5 *M* sodium carbonate, pH 10.0. The extent of hydrolysis was then determined from the fluorescence of the ionized product, 4-methylumbelliferone, as measured in a Fluorolog spectrofluorometer (SPEX, Edison, NJ) with an excitation wavelength at 380 nm and an emission wavelength at 455 nm. Preliminary experiments established that under these conditions 1% or less of the substrate was hydrolyzed in 30 min. Inhibition constants were then determined by means of Dixon plots of reciprocal velocity vs inhibitor concentration *(22)*.

Statistical Evaluation of Kinetically Determined Values

The statistical significances of differences observed among the means of triplicate determinations were evaluated by means of a one-tailed *t*-test

Table 1
X-ray Diffraction Data and Refinement Statistics for Cel5A Y245G

Unit cell $a = b =$	96.7
$c =$	258.6
Space group	$P3_221$
Unique reflections	52,895
Redundancy	6.3 (3.3)
Completeness (%)	90. (78.)
I/σ	10. (1.2)
$R_{mrgd\text{-}F}$ (%)[a]	9.9 (74.)
Average B-factor	
Main chain (Å^2)	17
Side chain	24
Resolution range (Å)	8–2.4
R-value (%)	19.7
Highest bin (2.4–2.5 Å)	42.2
R-free (%)[b]	23.4
Stereochemical ideality (root mean square deviations)	
Bonds (Å)	0.005
Angle distances (Å)	0.023
Distances from	
Restraint planes (Å)	0.026
Zero chiral volumes (Å^3)	0.029
Nonzero chiral volumes (Å^3)	0.035

[a]$\Sigma | <F_P> - <F_Q> | / 0.5\Sigma <F_P> + <F_Q>$, in which F_P and F_Q are subsets of reflections, and $<F_P>$ and $<F_Q>$ are the mean of each subset *(40)*. The values of the outermost resolution bins are shown inside parentheses.

[b]*R*-free was calculated using every tenth reflection (10% of the total data).

(23) using the percentage points of the *t*-distribution tabulated by Pearson and Hartley *(24)*.

Crystallography

The Y245G crystal grew in 2 mo from an 18-µL hanging drop containing 4.6 mg/mL of Cel5A, 2.2 *M* NaCl, 0.3 *M* mannoheptose, and 0.1 *M* sodium citrate buffer (pH 4.6) at room temperature. Crystals were also successfully grown in the presence of cellobiose and cellopentaose; however, they did not diffract as well as the crystals grown with mannoheptose. X-ray diffraction data for the Y245G were measured using an R-AXIS IV at a 2θ angle of 25° and crystal-to-detector distance of 250 mm. The CuK_α X-ray source was a Rigaku RU-H3RHB generator focused by Osmic-type, multilayer diffraction mirrors. Data reductions were carried out using SCALEPACK *(25)*. Cycles of manual model adjustment using InsightII and refinement of positional parameters and B-factors using SHELXL97 *(26)* resulted in the high-quality structure (Table 1). The two molecules in the asymmetric unit are virtually identical and were restrained as such during

the refinement. The resulting structure was verified using Procheck *(27)* and Whatif *(28)*.

Density Functional Calculations

DFT calculations *(18)* were carried out with the PQS 1.0 software from PQSLLC. The 6-311G** basis set *(29)* was used with Becke's three-parameter functional B3LYP *(30)* as deposited in GAUSSIAN94 *(31)*. To estimate the relative binding energies, calculations were performed for fragments of WT and Y245G, consisting of the segments 241–247 in the presence and absence of Glc1. Because the crystal structure of Y245G in complex with substrate is not known (we designed the mutant to bind less tightly), the position of Glc1 on Y245G was chosen to be identical to that of WT Glc1 in all interactions except those involving residues 241–247. The structures of WT cellotetraose and Y245G were then overlaid (omitting the 241–247 segments), and the Glc1 binding site was identified.

Owing to the time required to perform quantum chemical calculations on systems of the size involved in our study, it was not possible to select larger fragments, to consider the effects of solvation, or to optimize geometry. The effects of geometry optimization on relative energies can be considerable and cannot be predicted *a priori*. Thus, we consider the main value of these calculations their ability to confirm that the proposed interpretations of our observations (*see* above) are not unreasonable from first principles *(32)*. The binding energy for Glc1 on WT enzyme was found from

$$\Delta E_{\text{binding}} = E_{\text{wt–Glc1}} - E_{\text{wt}} - E_{\text{Glc1}}$$

in which $E_{\text{wt–Glc1}}$, E_{wt}, and E_{Glc1} are the total energies of the fragments of WT Glc1, WT, and Glc1, respectively. For Glc1 the full molecular structure was used. A corresponding equation was applied for the binding energy of Glc1 on Y245G, substituting $E_{\text{Y245G–Glc1}}$ for $E_{\text{wt–Glc1}}$. For the former, –6.6 kcal/mol was obtained, and –3.0 kcal/mol for the latter, yielding a binding energy difference of ~3.6 kcal/mol for Glc1 on WT and Y245G, favoring WT. Similar calculations were performed for the energy $E_{\text{Y245G/wt}}$ of WT in the conformation of Y245G at the site of mutation and the energy $E_{\text{wt/Y245G}}$ for Y245G in the conformation of WT at the site of mutation. The difference of $E_{\text{wt/Y245G}} - E_{\text{Y245G}}$ was found at 1.5 kcal/mol; the difference of $E_{\text{Y245G/wt}} - E_{\text{wt}}$ was found at >100 kcal/mol.

Results and Discussion

Catalytic Activities of Cel5A Mutants

The mutant Cel5A enzyme, Y245G, was generated by polymerase chain reaction mutagenesis, and the mutant and WT Cel5A catalytic domains were chromatographically purified. Purification of the native and mutant enzymes destined for kinetic analysis was crucial for this study, because specific

activities must be compared as precisely as possible. For this reason, the molar extinction coefficients of the recombinant enzymes were calculated by considering the specific change in amino acid composition.

As discussed in the Introduction, analysis of the X-ray crystallographic structure of the WT enzyme suggested that removal of the glucosyl-binding platform provided by Tyr245 would substantially decrease the affinity of the leaving-group binding site for cellobiose. The results of initial velocity kinetic experiments have shown that mutation of Tyr245 to glycine does indeed produce a large (more than 15-fold) decrease in affinity of the active site for cellobiose, in that the K_i for inhibition of the hydrolysis of 4 MUC by the WT enzyme is 1.88 ± 0.16 m*M*, but is increased to 29.7 ± 3.8 m*M* for the Y245G mutant. This increase in the value of K_i indicates a reduction in Cel5A/cellobiose binding energy on the order of 1.5 kcal/mol.

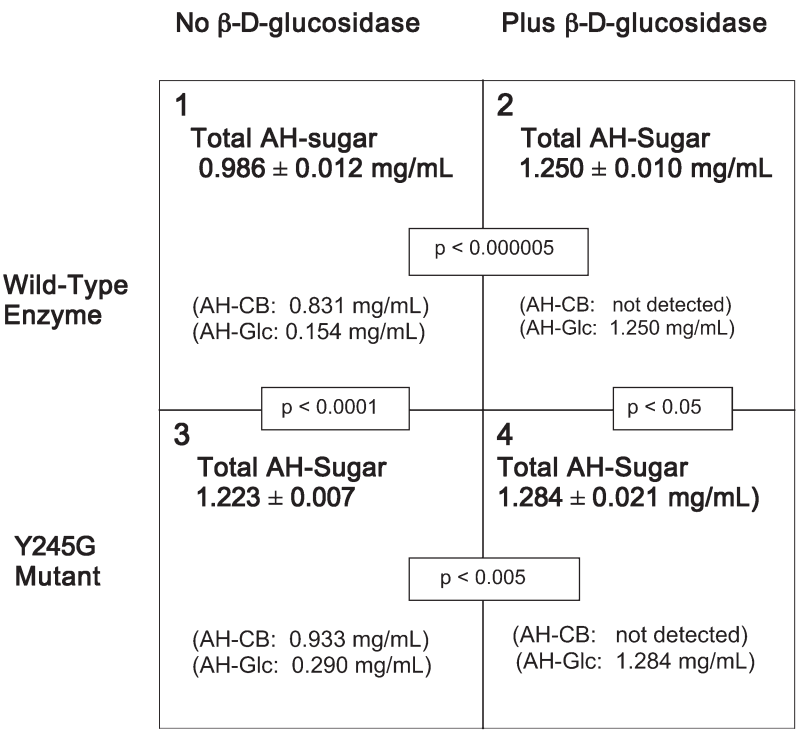
Closed-Tube Assays of Activity vs Swollen Cellulose

The decreased value of $K_{i,\text{cellobiose}}$ for the Y245G mutant is reflected in enhanced kinetics of hydrolysis of the polymeric substrate, PASC under conditions that lead to the accumulation of the product cellobiose in the digestion mixture (Table 2). PASC is a "good" substrate for endoglucanases in that a substantial portion of the crystalline cellulose has been converted into an amorphous, hydrated form sterically susceptible to attack by endoglucanases. The closed-tube PASC saccharification experiments reported in Table 2 were designed to involve solubilization of a significant fraction of this substrate and, thus, to result in the accumulation of significant concentrations of the principal products, cellobiose and glucose.

In the 2 × 2 experimental matrix in Table 2, the two conditions on the left (1 and 3) describe the saccharification results for the WT and Y245G mutant enzymes, respectively, each acting alone against PASC. As shown by the production of soluble "total anhydrosugars" (a method of reporting that makes the calculation of "percentage of substrate solubilized" independent of the eventual composition of the soluble-sugar mixture), the mutant enzyme solubilized almost 27% more of the PASC in 4 h than did the WT. The values shown for total anhydrosugars are averages of triplicate determinations. A one-tailed *t*-test using critical values tabulated by Pearson and Hartley *(24)* indicates that the measured difference between soluble-sugar production by the mutant and WT enzymes (conditions 3 and 1, respectively) is significant at the $p < 0.0001$ level, as shown in the small box inset in Table 2.

The data in the two right-hand cells of the matrix in Table 2 provide a very strong suggestion concerning the mechanism by which the Tyr → Gly substitution results in the enhancement of hydrolysis kinetics seen in the mutant. Conditions 3 and 4 are identical, respectively, to conditions 1 and 3, except that in conditions 2 and 4, sufficient purified *A. niger* β-D-glucosidase was added to convert the cellobiose released by the endoglucanase almost quantitatively into glucose, thus driving the cellobiose

Table 2
Release of Soluble Sugars from PASC by WT and Mutant Cel5A Enzymes, in Presence and Absence of *A. niger* β-D-Glucosidase[a]



[a]Assays were carried out at 38°C, pH 5.0 in 20 m*M* acetate in closed vessels. Substrate loading was 5 mg/mL; Cel5A loadings were 28 μg (approx 70 nanomol)/mL. Purified *A. niger* β-D-glucosidase (where present) was added at 45 μg/mL. AH-CB anhydrocellobiose; AH-Glc, anhydroglucose.

concentration to undetectably low levels. When β-D-glucosidase alone was added to PASC under identical conditions and loadings, the apparent sugar release (as glucose) was <2% of the smallest total AH-sugar release seen under any of the conditions in the matrix, and too small to be reliably quantified. It would appear likely, therefore, that the only significant effect of the added β-D-glucosidase on the sugar release values in conditions 2 and 4 is the hydrolysis to glucose of the cellobiose released from PASC by the WT and mutant endoglucanases. In the absence of accumulated cellobiose, the difference between the activities of the mutant and WT enzymes essentially disappeared (conditions 3 and 4 of Table 2).

When the values of $K_{i,\text{cellobiose}}$ are expressed in terms of milligrams of "anhydrocellobiose" per milliliter, the numbers are 0.611 mg/mL for the WT enzyme and 9.66 mg/mL for the mutant. The cellobiose accumulations in conditions 1 and 3 (digestions in the absence of β-D-glucosidase) are thus seen to be approx 1.36 times K_i for the WT enzyme, but <0.1 times

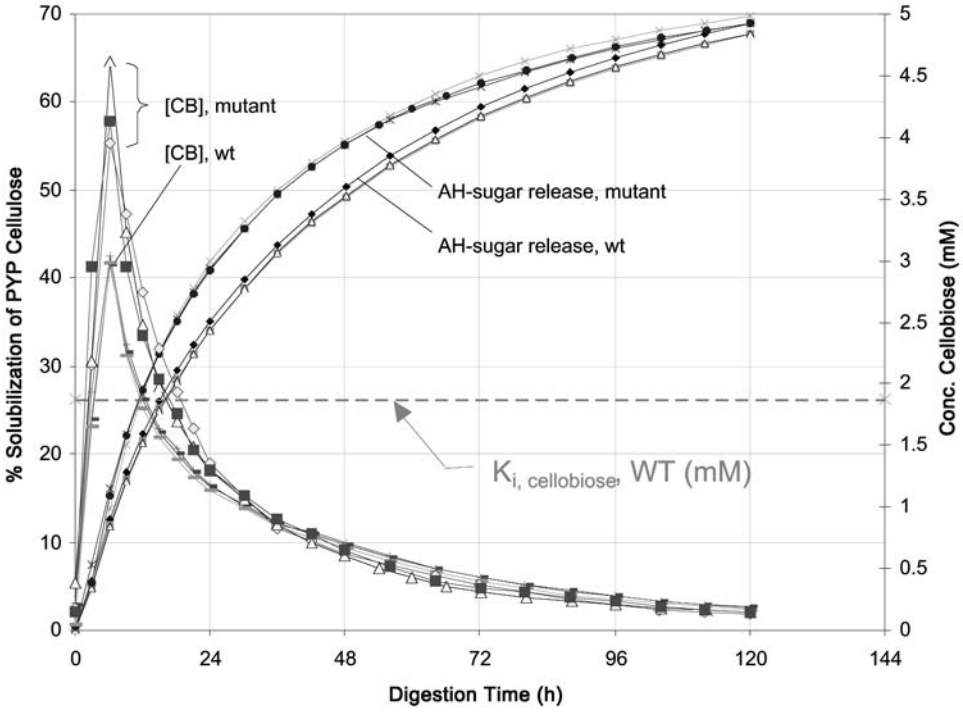


Fig. 3. Effect of product cellobiose (CB) concentrations on kinetics of saccharification of PYP by WT and Y245G mutant versions of Cel5A (assayed in combination with *T. reesei* Cel7A, as described in Materials and Methods). The concentrations of cellobiose in the DSA effluent fractions (left-hand axis) are coplotted with the saccharification progress curves (cumulative sugar released, as a percentage of that theoretically available) for the binary mixtures (1:19 molar ratio) of endoglucanase (Cel5A-WT or Cel5A-Y245G) with *T. reesei* Cel7A. The horizontal dashed line at 1.88 m*M* cellobiose represents the value of K_i for inhibition of the WT Cel5A by cellobiose; the corresponding K_i value for the mutant Y245G, at 29.7 m*M*, is far off the scale of the plot.

K_i for the mutant. Given that the difference in activity between mutant and WT enzyme largely disappears when accumulation of cellobiose is prevented by conversion of the cellobiose into glucose as in conditions 2 and 4, the "original" activity difference (conditions 1 and 3) between mutant and WT can be assigned, with a reasonable degree of confidence, to the differences in susceptibility to cellobiose inhibition. The only product inhibition occurring in the assays of conditions 2 and 4 is from glucose. Given that glucose is much less inhibitory than cellobiose ($K_{i,\text{glucose}}$ for the WT enzyme is approx 360–370 m*M*, or 59.5 mg/mL in "anhydroglucose" terms), inhibition by any of the glucose accumulations found in these experiments is unlikely to be a significant issue.

The implications of the experiments summarized in Fig. 3 are two fold: First, the presence in the reaction mixture of sufficient β-D-glucosidase to prevent accumulation of cellobiose tends to eliminate most of the effect of the Y245G mutation by bringing the performance of the WT enzyme almost

up to the level of the mutant under these conditions. Second, the presence of the Y245G mutation in the endoglucanase tends to eliminate most of the effect of adding β-D-glucosidase, in that the mutant enzyme in the *absence* of β-D-glucosidase performs almost as well as in the presence of β-D-glucosidase. This rough interchangeability of the effects of Y245G mutation and of the addition of β-D-glucosidase might be said, in somewhat colloquial terms, to mean that we have "engineered in" a β-D-glucosidase effect.

Membrane-Reactor Assays of Activity vs Biomass Cellulose

The PASC-saccharification experiments discussed may be considered to mimic, in a limited way, one industrial application of cellulase enzymes, namely the saccharification step of separate saccharification and fermentation (SHF), which is one possible configuration for the process of converting biomass cellulose into fuel ethanol, or other chemical products. Although a highly processed pure cellulose such as PASC is unlikely to be chosen as an industrial feedstock, the PASC experiments do resemble SHF in that the cellulose depolymerization is carried out in a closed system, so that products accumulate to substantial concentrations. The discussion of the Y245G mutant endoglucanase and its catalytic performance is now concluded with an example of additional experiments carried out under conditions mimicking those encountered in the processing of a candidate cellulosic feedstock in a related industrial application (i.e., simultaneous saccharification and fermentation [SSF]).

The progress curves presented in Fig. 3 illustrate the enzymatic saccharification of the cellulose component of dilute-acid-pretreated yellow poplar (PYP), which is poplar sawdust from which most of the hemicellulosic material, and some of the lignin, has been removed by dilute-acid hydrolysis at high temperature. Left behind in the PYP substrate is a sterically complex mechanical intermixture of predominantly cellulose (approx 58%) and lignin (approx 35%), arranged in a matrix retaining substantial elements of the original wood structure. Attack on this physically and chemically heterogeneous substrate was carried out (Fig. 3) by binary enzyme mixtures formed by adding either WT or Y245G mutant Cel5A endoglucanase to *T. reesei* CBH-I (Cel7A). In nature, and in virtually any industrial process involving substantial enzymatic saccharification of biomass material, the effective digestion of cellulose is carried out not by one enzyme, but by mixtures of cellulolytic enzymes acting synergistically *(8,33)*. In the experiments shown in Fig. 3, the binary mixture of one endoglucanase (Cel5A WT or the Y245G mutant) with one exoglucanase (Cel7A) can be regarded as a minimal effective system for attack on an insoluble, and still significantly crystalline, cellulosic material.

The digestions in Fig. 3 were carried out in a stirred membrane reactor set up so that the products of the digestion (small molecular weight solubilized sugars) were continually swept out of the reactor by the buffer flux, which was then collected in timed fractions and analyzed for sugar content

to provide the cumulative sugar production progress curves shown in Fig. 3. In this assay (*see* Materials and Methods; *[20]*), removal of the solubilized sugars by the buffer flux mimics the consumption of sugars by fermentative organisms in SSF. In both cases (DSA and SSF), continuous removal of sugars greatly reduces product inhibition of the cellulases by driving the pseudo-steady-state concentration of the sugars to a much lower level than would be present were the sugars allowed to accumulate without removal, but does not reduce the sugar concentrations to zero. This last point will be revisited later.

The progress curves in Fig. 3 show enhanced initial kinetics for the hydrolysis of the cellulose component of PYP by the binary enzyme mixture of the Y245G mutant and Cel7A, relative to the performance of an otherwise identical mixture formulated using the WT endoglucanase. The initial slope of the progress curve for the enzyme mixture containing the mutant is significantly steeper than that for the enzyme mixture containing the WT endoglucanase, but over the course of the reaction the rate of conversion for the mutant-containing mixture decreased more rapidly than did the rate for the mixture containing WT endoglucanase. By the 120-h point in the digestion, the WT mixture had essentially caught up in terms of cumulative sugar release, showing (in the averages for triplicate determinations) 98.8% as much cumulative sugar release as the "mutant" mixture. As shown by the coplots of the effluent cellobiose concentration with the overall cumulative progress curves (Fig. 3), the mutant-containing mixture gained most of its advantage in the very early stage of the reaction, where the liquid-phase cellobiose concentrations were at their highest, spiking to 1.59 times the endoglucanase K_i for the WT mixture, but to only 0.14 times the (much higher) K_i value for the mutant-containing mixture.

The fact that a single point mutation in a single enzyme has such a clear effect in this case, even though the mutated enzyme is the minority component (5% on a molar basis) in the enzyme mixture, is very probably related to the synergistic action of endoglucanases and exoglucanases in the depolymerization of insoluble cellulose. In addition to being able to release soluble sugars from cellulose themselves by successive attacks, the endoglucanases play an important role by potentiating the action of the exoglucanases. The effect of a mutation in the minority enzyme component is therefore amplified by an effect on the activity of the majority component.

The extent to which the relief of product inhibition by the Y245G mutation in the endoglucanase increases the activity of the cellulase mixture can be quantitated, both conveniently and meaningfully, by "time-to-target" analysis of the two progress curves in Fig. 3. In this approach, the continuous progress curves are used to generate estimates of the time required for the two enzyme mixtures to accomplish the *same* extent of conversion of the substrate. The reciprocals of these "times to target" are then used as measures of relative activities for the two mixtures. For example, an enzyme or enzyme mixture that converts 30% of the substrate in 1 h is

regarded as having twice the activity of an enzyme or mixture that accomplishes the same thing in 2 h, or *twice* the time *(34)*. This approach is especially attractive in dealing with substantial conversion of heterogeneous substrates such as PYP, because, even though the nature of the substrate changes over the course of a substantial conversion, it is not unreasonable to assume that two enzymes or mixtures that convert the substrate to the *same extent* will have acted on substrate of essentially the same nature over the course of the reaction. Using the "reciprocal time to target" as an estimator of relative activity, we find that the ratio of rates is maximal when a value near 35% is chosen as the target extent of digestion. The mutant-containing mixture is found to reach 35% conversion in 17.7 ± 0.3 h (average of triplicate determinations), whereas the enzyme mixture containing the WT endoglucanase requires 24.7 ± 0.2 h to accomplish the same extent of conversion. These digestion times correspond to the reciprocals 0.0405 ± 0.0004 h^{-1} (WT) and 0.0565 ± 0.0005 h^{-1} (mutant). The difference between these two means is significant at the $p < 0.0001$ level, and the ratio of the means (mutant/WT) is 1.396, indicating that in this substantial conversion of the cellulose content of a realistic industrial biomass feedstock, the mixture containing the mutant endoglucanase exhibits almost 40% greater activity than the mixture utilizing the WT endoglucanase.

Structural Assessment for Y245G Enhanced Activity

Overall structural variations between WT *(10)* and Y245G are minimal. The root mean square deviation of C_α between WT and Y245G is 0.22 Å. Even though the overall structures were similar, important structural changes occurred in mutant Y245G at site 246, but not at site 245. That is, compared to WT, the ψ torsional angle of residue 246 in the crystal structure of Y245G is shifted from 67.6 to 142.5°. In this state, the carbonyl group of Pro246 is positioned to the inside of the catalytic cleft and readily available for hydrogen bonding with water. The water molecule is not in a position in which it can form a hydrogen bond with the hydroxyl groups of Glc1 (Fig. 4). A further consequence of the torsional change at Pro246 is the retraction of Gln247 away from the enzyme cavity. In the WT-substrate complex, Nε2 of Gln247 interacts with O2 of Glc1. Thus, one may notice that the mutation of WT to Y245G reduces the binding energy between the leaving group and the enzyme by two means: by removing a hydrophobic platform residue and by lengthening a hydrogen bond by ~0.5 Å. The experimental estimate for the reduction in binding energy (~1.5 kcal/mol) is supported by the DFT calculations (*see* Density Functional Conditions), which yield a value of ~3 kcal/mol.

The DFT calculations are in agreement with the assumption that the main-chain torsional angle change at Pro246 between WT and Y245G is caused by both torsional strain and steric interactions. Torsional strain is indicated by the fact that the density function energy of Y245G calculated with the ϕ,ψ-torsional angles of WT at site 246 is ~1.5 kcal/mol higher than

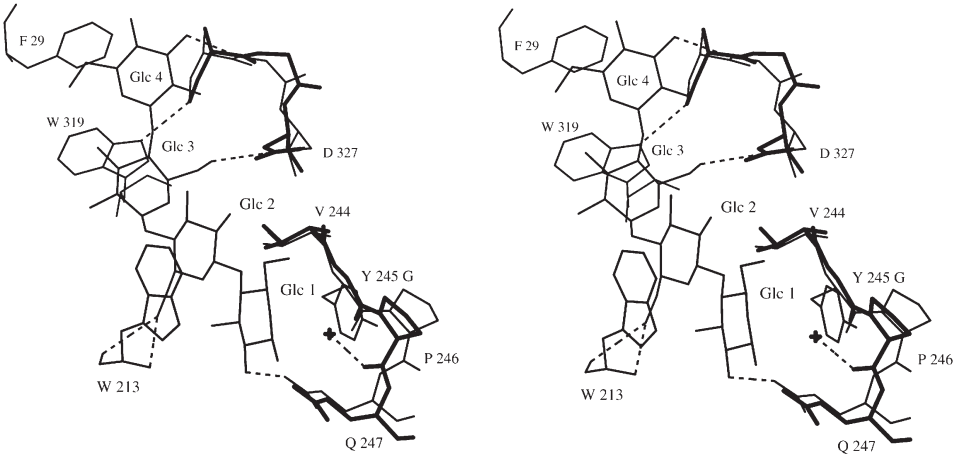


Fig. 4. Stereoscopic view of effects of Y245G mutation. Equivalent loops in Y245G (heavy lines) and the WT (thin line) are shown. Y245G mutant was grown in the presence of mannoheptose; however, no evidence for the bound ligand was observed. The four glucosyl units shown are from the WT-substrate complex *(10)*. Dashed lines show hydrogen-bonding interactions of <3.2 Å.

that of Y245G with site 246 in the crystal structure. This finding is also in agreement with the fact that the ϕ,ψ-torsional angles of WT at Pro246 (–68.0°, 67.6°) are in a scarcely populated region of the Ramachandran plot for Pro residues, whereas the values in Y245G, at –78.7 and 142.5°, are commonly observed. The importance of steric interactions is apparent from the fact that if WT were to adopt the ψ angle found at Pro246 in the crystal structure of Y245G, then O-Gln247 and Cδ2-Tyr245 would be subject to a highly unfavorable interaction at ~2.4 Å. DFT calculations indicate that the corresponding destabilization can be on the order of several tens of kilocalories per mole. In other proteins, a small but significant fraction of non-Gly residues have been found to adopt ϕ,ψ angles that are energetically unfavorable *(35)*. The mutation of those residues in a model protein, *Staphylococcal* nuclease, showed that relieving such strain energy could increase the stability of the protein by 1 to 2 kcal/mol with respect to the WT *(36)*. The use of DFT results agreed well with the labor-intensive, experimental determination of strain energy in the model protein while reducing the time and labor required. Such success supports the general use of DFT in protein engineering.

To address the question of whether the loop in Y245G had become flexible by the loss of the side chain of Tyr to Gly, we performed a detailed analysis of the temperature factors. Relative B-factor values within a molecule have been shown to contain some information about thermal atomic displacements *(37–39)*. After comparing the relative B-factor values of WT and Y245G near their binding sites, we concluded that the mutant Gly did not significantly increase the thermal displacements. By contrast,

the WT cellotetraose complex exhibited significantly higher temperature factors, perhaps because cellotetraose is a true substrate, and the enzyme is in a superposition of four different states *(10)*.

In conclusion, we propose that the enhanced catalytic activity of the endocellulase Cel5A mutant (Y245G) is primarily owing to a reduction in product inhibition. The present results do not rule out the possibility that part of the total "inhibition" that is relieved may actually reflect reversal of the depolymerization reaction by attack of bulk solution cellobiose on the glycosyl enzyme (i.e., transglycosylation). Whether the relieved inhibition is attributed to one or the other or to a combination of both of these mechanisms, it is important to note that both mechanisms involve recombination of released product with the enzyme-active site. Thus, the central messages of this study are as follows: Theoretical binding-energy calculations utilizing high-resolution X-ray crystallographic structures of Cel5A indicated that a specific mutation (Tyr245 to Gly245) should reduce the affinity of the enzyme-active site for the product, cellobiose. In addition, initial velocity enzyme kinetic measurements on both the native enzyme and the mutant revealed that the affinity for cellobiose in the mutant was indeed reduced substantially when compared with the original enzyme (K_i value 15.8-fold larger in the mutant). Moreover, in further kinetic studies involving substantial conversion of two different insoluble cellulosic substrates (one a feasible industrial biomass feedstock) under simulated industrial process conditions, the reduced susceptibility of the engineered enzyme to cellobiose inhibition was shown, as also predicted, to translate into enhanced rates of depolymerization of cellulose. These combined results are therefore a powerful confirmation of the value of an information-based approach, using structural and kinetic data to drive site-directed mutagenesis, in engineering enzymes for specific applications.

Acknowledgments

This work was funded in part by the DOE Office of the Biomass Program. We greatly appreciate Lothar Schäfer for his critical reading and computer usage, the Barry Goldwater Foundation, and a POLYED Summer Scholarship.

References

1. Wyman, C. E., Bain, R. L., Hinman, N. D., and Stevens, D. J. (1993), in *Renewable Energy: Sources for Fuels and Electricity*, Johansson, T. B., Kelly, H., Reddy, A. K. N., and Williams, R. H., eds., Island Press, Washington, DC.
2. Sheehan, J. J. (1994), in *Enzymatic Conversion of Biomass for Fuels Production,* ACS Series 566, Himmel, M. E., Baker, J. O., and Overend, R. P., eds., American Chemical Society, Washington, DC, pp. 1–52.
3. Bergeron, P. (1996), in *Handbook on Bioethanol*, Wyman, C. E., ed., Taylor & Francis, Washington, DC, pp. 61–88.

4. Hsu, T.-A. (1996), in *Handbook on Bioethanol*, Wyman, C. E. ed., Taylor & Francis, Washington, DC, pp. 179–195.
5. Himmel, M. E., Adney, W. S., Baker, J. O., Elander, R., McMillan, J. D., Nieves, R. A., Sheehan, J. J., Thomas, S. R., Vinzant, T. B., and Zhang, M. (1997), in *Fuels and Chemicals from Biomass*, ACS Series 666, American Chemical Society, Washington, DC, pp. 2–45.
6. Himmel, M. E., Adney, W. S., Grohmann, K., and Tucker, M. P. (1994), US patent 5,275,944.
7. Divne, C., Stahlberg, J., Reinikainen, T., Ruohonen, L., Pettersson, G., Knowles, J. K. C., Teeri, T. T., and Jones, A. (1994), *Science* **265,** 524–528.
8. Baker, J. O., Adney, W. S., Thomas, S. R., Nieves, R. A., Chou, Y.-C., Vinzant, T. B., Tucker, M. P., Laymon, R. A., and Himmel, M. E. (1995), in *Enzymatic Degradation of Insoluble Carbohydrates*, ACS Series 618, Saddler, J. N., and Penner, M. H., eds., American Chemical Society, Washington, DC, pp. 113–141.
9. Irwin, D. C., Spezio, M., Walker, L. P., and Wilson, D. B. (1993), *Biotechnol. Bioeng*. **42,** 1002–1013.
10. Sakon, J., Adney, W. S., Himmel, M. E., Thomas, S. R., and Karplus, P. A. (1996), *Biochemistry* **35,** 10,648–10,660.
11. Koshland, D. E. J. (1953), *Biol. Rev*. **28,** 416–436.
12. McCarter, J. D. and Withers, S. G. (1994), *Curr. Opin. Struct. Biol*. **4,** 885–892.
13. Sinnott, M. L. (1990), *Chem. Rev*. **90,** 1171–1202.
14. Rouvinen, J., Bergfors, T., Teeri, T., Knowles, J. K. C., and Jones, T. A. (1990), *Science* **249,** 380–386.
15. Spezio, M., Wilson, D. B., and Karplus, P. A. (1993), *Biochemistry* **32,** 9906–9916.
16. Sakon, J., Irwin, D., Wilson, D. B., and Karplus, P. A. (1997), *Nat. Struct. Biol*. **4,** 810–818.
17. Moreau, A., Shareck, F., Kluepfel, D., and Morosoli, R. (1994), *Eur. J. Biochem*. **219,** 261–266.
18. Parr, R. G. and Yang, W. (1989), *Density-Functional Theory of Atoms and Molecules*, Oxford University Press, Oxford, UK.
19. Baker, J. O., Ehrman, C. I., Adney, W. S., Thomas, S. R., and Himmel, M. E. (1998), *Appl. Biochem. Biotechnol*. **70,** 395–403.
20. Baker, J. O., Vinzant, T. B., Ehrman, C. I., Adney, W. S., and Himmel, M. E. (1997), *Appl. Biochem. Biotechnol*. **63–65,** 585–595.
21. Thomas, S. R., Laymon, R. A., Chou, Y.-C., Tucker, M. P., Vinzant, T. B., Adney, W. S., Baker, J. O., Nieves, R. A., Mielenz, J. R., and Himmel, M. E. (1995), in *Enzymatic Degradation of Insoluble Polysaccharides*, ACS Series 618, Saddler, J. N. and Penner, M. H., eds., American Chemical Society, Washington, DC, pp. 208–236.
22. Segel, I. H. (1975), in *Enzyme Kinetics*, John Wiley & Sons, New York, pp. 109–111.
23. Snedecor, G. W. and Cochran, W. G. (1967), *Statistical Methods,* Iowa State University Press, Ames.
24. Pearson, E. S. and Hartley, H. O. (1970), *Biometrika Tables for Statisticians*, Cambridge University Press, Cambridge, UK.
25. Otwinowski, Z. (1993), in *Proceedings of the CCP4 Study Weekend: Data Collection and Processing,* SERC Daresbury Laboratory, Warrington, UK.
26. Sheldrick, G. M. (1990), *Acta Crystallogr. Sect. A* **46,** 467–473.
27. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993), *J. Appl. Crystallogr*. **26,** 283–291.
28. Vriend, G. (1990), *J. Mol. Graph*. **8,** 52–56.
29. Krishnan, R., Binkley, J. S., Seeger, R., and Pople, J. A. (1980), *J. Chem. Phys*. **72,** 650–654.
30. Becke, A. D. (1993), *J. Chem. Phys*. **98,** 5648–5652.
31. Frisch, M. J. (1995), *Gaussian 94*, Gaussian, Pittsburgh, P A.
32. Ramek, M., Momany, F. A., Miller, D. M., and Schäfer, L. (1996), *J. Mol. Struct*. **375,** 189–191.
33. Nidetzky, B., Steiner, W., and Claeyssens, M. (1994), *Biochem. J*. **303,** 817–823.

34. Ghose, T. K. (1987), *Pure Appl. Chem*. **59,** 257–268.
35. Karplus, P. A. (1996), *Protein Sci*. **5,** 1406–1420.
36. Stites, W. E., Meeker, A. K., and Shortle, D. (1994), *J. Mol. Biol*. **235,** 27–32.
37. Kuriyan, J. and Weis, W. I. (1991), *Proc. Natl. Acad. Sci. USA* **88,** 2773–2777.
38. Ringe, D. and Petsko, G. A. (1986), *Methods Enzymol*. **131,** 389–433.
39. Stroud, R. M. and Fauman, E. B. (1995), *Protein Sci*. **4,** 2392–2404.
40. Diederichs, K. and Karplus, P. A. (1997), *Nat. Struct. Biol*. **4,** 269–275.